# Построение инвариантов на множестве временных рядов путем динамической свертки свободной переменной

***Стрижов В. В., Пташко Г. О.***
Москва, Вычислительный центр РАН
strijov@ccas.ru

Дано множество временных рядов, на котором заданы классы эквивалентности. Требуется построить модель — параметрическое семейство преобразований, которая задает инвариант, отображающий временные ряды в соответствующие классы. Модель выбирается из множества суперпозиций гладких монотонных функций. Предложенный алгоритм основан на методе поиска оптимальной регрессионной модели как произвольной суперпозиции порождающих функций [1].

Ранее были предложены методы классификации временных рядов с использованием динамической свертки времени [2]. Классификация выполнялась с учетом порогового значения пути наименьшей стоимости. Также классификация может быть выполнена с использованием пространства параметров функций, аппроксимирующих временные ряды [3] в качестве входной информации.

Особенностью предложенного алгоритма является то, что аппроксимация выполняется не на множестве временных рядов, а на множестве их путей минимальной стоимости, что позволяет сократить множество моделей-претендентов и уменьшить их сложность.

## Постановка задачи

Дано $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ — множество временных рядов, $\mathbf{x}_i = \{x_{it}\}_{t=1}^{T}$. На упорядоченных парах индексов $(i, j)$ временных рядов $(\mathbf{x}_i, \mathbf{x}_j)$ задано отношение эквивалентности $[\mathbf{x}]$, которое интерпретируется как принадлежность временных рядов к некоторому классу.

Задано множество монотонных параметрических гладких функций $\big\{g(\mathbf{b}, \cdot, \ldots, \cdot) \mid g\colon \mathbb{R} \times \cdots \times \mathbb{R} \to \mathbb{R}\big\}$, которое порождает множество суперпозиций $\{f_k(\mathbf{w}_k)\colon t \to t\}$ путем обобщенного индуктивного определения. Вектор параметров $\mathbf{w}$ есть присоединенные векторы $\langle \mathbf{b}_1, \ldots, \mathbf{b}_{r(k)} \rangle$ параметров функций, входящих в суперпозицию; размерность вектора $\mathbf{w}$ зависит от структуры суперпозиции: $\mathbf{w}_k \in \mathbb{R}^{W(k)}$.

Требуется выбрать такую модель $f_k(\mathbf{w})$ и найти такое множество допустимых значений ее параметров $[w_k] \subset \mathbb{R}^{W(k)}$, которые задают требуемый класс эквивалентности $[\mathbf{x}]$ при условии, что множество $[w_k]$ доставляет максимум целевой функции на всех парах временных рядов из $[\mathbf{x}]$.

### Аппроксимация пути наименьшей стоимости

Построим матрицу расстояний $\Omega(\mathbf{x}_i, \mathbf{x}_j) = \{\omega_{pq} = \|x_{ip} - x_{jq}\|_2\}_{p,q=1}^T$ между всеми парами элементов $(x_{ip}, x_{jq})$ временных рядов $\mathbf{x}_i, \mathbf{x}_j$.

Для отыскания пути наименьшей стоимости рассмотрим в матрице $\Omega$ все пути $\mathbf{s} = \{s_1, \ldots, s_C\}$ такие, что $s_1 = \omega_{11}$, $s_C = \omega_{T,T}$ и для произвольного $s_c = \omega_{pq}$, где $c = 1, \ldots, C-1$ значение $s_{c+1} = \omega_{p+u,q+v}$, где $u + v \in \{1, 2\}$. Вычислим стоимость пути $\mathbf{s} = C^{-1}\bigl(\sum_{c=1}^{C} s_c^2\bigr)^{1/2}$. Обозначим $\bar{\mathbf{s}}_{ij}$ — путь наименьшей стоимости для пары временных рядов $(\mathbf{x}_i, \mathbf{x}_j)$. Подробнее об алгоритме динамической свертки временных рядов см. [4].

Найдем такие параметры $\mathbf{w}_k(i,j)$, при которых модель наилучшим образом приближает путь $\bar{\mathbf{s}}_{ij}$. Оптимальные значения параметров заданы как $\bar{\mathbf{w}}_k = \arg\min S\bigl(f_k(\mathbf{w}_k(i,j)), \bar{\mathbf{s}}_{ij}\bigr)$.

В данной работе предлагается выбирать такую модель, которая не нарушает отношения эквивалентности $[\mathbf{x}]$, и для которой максимальное значение $S$ по всем парам $(i,j)$, таким, что $(\mathbf{x}_i, \mathbf{x}_j) \in [\mathbf{x}]$ минимально: $\max\limits_{i,j} S\bigl(f_k(\bar{\mathbf{w}}_k(i,j)), \bar{\mathbf{s}}_{ij}\bigr) \to \min$.

### Классы эквивалентности в пространстве параметров

Обозначим $\bar{f}_{ij}$ фиксированную по $k$ функцию $f_k(\bar{\mathbf{w}}_k(i,j))$. Множество последовательных отображений $\bar{f}_{12}, \bar{f}_{23}, \ldots, \bar{f}_{(\ell-1)\ell}$ задано множеством фиксированных параметров $\bar{\mathbf{w}}_{12}, \bar{\mathbf{w}}_{23}, \ldots, \bar{\mathbf{w}}_{(\ell-1)\ell}$, $\ell = N(N-1)/2$, полученных в результате идентификации.

Элементы множества $\{\bar{f}_{ij}\}$, сохраняющие класс эквивалентности $[\mathbf{x}]$, задают группу преобразований $G$. Для произвольной тройки $i, j, h$ индексов временных рядов из $[\mathbf{x}]$ справедлива коммутативная диаграмма

$$\begin{tikzcd} \mathbf{x}_i \arrow[rr, "f_{ij}"] \arrow[dr, "f_{ih}"'] & & \mathbf{x}_j \\ & \mathbf{x}_h \arrow[ur, "f_{hj}"'] & \end{tikzcd} \quad \text{и} \quad G = \begin{pmatrix} \mathrm{id} & f_{12} & \cdots & f_{1N} \\ f_{12}^{-1} & \mathrm{id} & \cdots & f_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{1N}^{-1} & f_{2N}^{-1} & \cdots & \mathrm{id} \end{pmatrix}.$$

Группа $G$ задает классы эквивалентности в пространстве параметров, $[\mathbf{w}] \ni \{\bar{\mathbf{w}}_k(i,j)\}$. Рассмотрим такие модели $f_k$, которые позволяют доопределить $[\mathbf{w}]$ до выпуклой оболочки $[w_k]$ элементов этого класса без потери отношения эквивалентности. Такое доопределение позволит классифицировать временные ряды, не включенные во множество $[\mathbf{x}]$ ранее.

Для каждого нового временного ряда $\mathbf{x}_{N+1} \notin X$ заданной ранее размерности находим $\{\bar{\mathbf{w}}_k(i,N)\}$ для $i \in \{1, \ldots, N\}$. Ряд $\mathbf{x}_{N+1}$ принадлежит $[\mathbf{x}]$, если вектор параметров $\bar{\mathbf{w}}_k(i, N+1)$ принадлежит $[w_k]$. Таким образом, по построенному классу эквивалентности $[w_k]$ можно определить принадлежность нового временного ряда к $[\mathbf{x}]$.

### Заключение

Предложенный метод был использован для классификации временных рядов давления в камере внутреннего сгорания дизельного двигателя. Непосредственное вычисление значения пути оптимальной стоимости не позволило решить задачу классификации исследуемых временных рядов, так как стоимость двух несовпадающих путей временных рядов из разных классов часто оказывалась одинаковой. Создание моделей, аппроксимирующих эквивалентные временные ряды, также не позволило решить данную задачу, так как классификацию при этом приходилось выполнять в пространстве параметров, которое имело большую размерность. Предложенный метод позволил

разделить данные временные ряды на кластеры, так как классификация выполняется в пространстве параметров небольшой размерности.

Работа выполнена при поддержке РФФИ, проект № 07-07-00181.

### Литература


[1] *Стрижов В. В., Пташко Г. О.* Алгоритмы поиска суперпозиций при выборе оптимальных регрессионных моделей. — М: ВЦ РАН, 2006. — 42 с.

[2] *Pavlidis I., Singh R., Papanikolopoulos N.* Recognition of On-Line Handwritten Patterns Through Shape Metamorphosis. // Proceedings of the 13th International Conference on Pattern Recognition. — 1996. — Vol. 3. — Pp. 18–22.

[3] *Стрижов В. В.* Поиск параметрической регрессионной модели в индуктивно заданном множестве // Вычислительные технологии. — 2007. — Т. 12, № 1. — С. 93–102.

[4] *Keogh E. J., Pazzani M. J.* Derivative Dynamic Time Warping // First SIAM International Conference on Data Mining (SDM'2001), Chicago, USA. 2001. — `http://www.ics.uci.edu/~pazzani/Publications/sdm01.pdf`.